REMARKS

This amendment responds to the office action mailed July 8, 2003. In the office action the Examiner:

- rejected claims 1-19, and 21 under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 4,992,940 to Dworkin (hereinafter "Dworkin"); and

- rejected claim 20 under 35 U.S.C. 103(a) as being unpatentable over Dworkin in view of U.S. Patent No. 6,490,567) to Gregory (hereinafter "Gregory").

After entry of this amendment, the pending claims are: claims 1-21, and 23-34. Applicants respectfully traverse the rejection of claims 1-21. With this amendment claims 1-4, 7-8, 10, 11, and 14-21 have been amended for clarity and new claims 23-34 have been added to more particularly claim certain aspects of the invention. No new matter has been added.

THE 35 U.S.C. 103(a) REJECTIONS SHOULD BE WITHDRAWN

With this amendment, each of the independent claims has been amended to indicate the extraction pattern claimed in original claim 3. For the reasons explained below, the pending claims, as amended, are patentable over Dworkin and Gregory, either alone or in combination. The pending claims each indicate, or depend from a claim that indicates, (i) the development of an extraction pattern, based on the output of a target web site, that extracts data from the target web site and (ii) receiving a value that can be used as an extraction parameter for the developed extraction pattern.

Exemplary development of an extraction pattern

An example of the development of an extraction pattern based on the output of a target web site (*e.g.*, a web page) is provided on page 25 of the specification in conjunction with Figure 16 where, as indicated in new claims 23, 27, 28 and 33, the extraction pattern comprises a pre-condition regular expression (1606), a portion of data of interest regular expression (1608), and a post-condition regular expression (1610):

$$\left[(\text{<dt>}) \right] \left[(\text{A Prayer for Owen Meany}) \right] \left[\text{}) \right].$$

In this example, pre-condition 1606, regular expression 1608 and post-condition regular expression 1610 are separated by large brackets for clarity. The above-identified HTML code represents a portion of the output of a target web site (*e.g.*, a web page) when a user requested information from the target web site for the book entitled "A prayer for Owen Meany."

In the example provided by Fig. 16, development of an extraction pattern, based on the output of a target web site, involves replacing the portions of the extraction pattern that are unique to the title "A Prayer for Owen Meany" with regular expressions that will match other titles. This replacement is illustrated in Fig. 17, where specific references to the book "A prayer for Owen Meany" have been replaced with wildcards:

$$\left[(\text{<dt>}) \right] \left[(.*?) \right] \left[\text{}) \right]$$

Thus, the extraction pattern of Fig. 17 represents a development of the extraction pattern of Fig. 16. The developed extraction pattern of Fig. 17 is no longer limited to extracting information about the book "A Prayer for Owen Meany." The extraction pattern can now be passed an extraction parameter, in this case, the title of a book for which a user is interested. As indicated in each of Applicants' independent claims, a value that can be used as an extraction parameter for the developed extraction pattern is used to query a web site for data of interest (*e.g.*, product information, news, *etc.*). For example, in the case of the extraction pattern of Fig. 17, the extraction parameter is a book title and the value is a specific book title (*e.g.*, The Catcher in the Rye).

Dworkin does not teach an extraction pattern

On page four of the July 8, 2003, Office Action, the Examiner indicated that Dworkin teaches extraction patterns based on the disclosure in Dworkin of a database that holds the equivalent of thousands of catalogs of individual suppliers. This is not the case. Applicants'

extraction patterns cannot be equated to Dworkin's database. Applicants' extraction patterns parse the output of target web sites. As such, they parse the output looking for matching subsequences. Furthermore, in order to work, Applicants' extraction patterns are passed a value that can serve as an extraction parameter for the extraction pattern. For example, the developed extraction pattern of Fig. 17:

$$\left[(<\text{dt}><\text{b}><\text{a href="/exec/obidos/ASIN/.*?">}\right]\left[(.*?)\right]\left[\ </\text{a}></\text{b}>)\right]$$

is passed a value that can serve as an extraction parameter. Consider the case in which the value "The Catcher in the Rye" is provided as an extraction parameter. The value for the extraction parameter is then inserted in the extraction pattern, in accordance with the claimed invention, as follows:

$$\left[(<\text{dt}><\text{b}><\text{a href="/exec/obidos/ASIN/.*?">}\right]\left[(\text{The Catcher in the Rye})\right]\left[\ </\text{a}></\text{b}>)\right]$$

This populated extraction pattern is then used to parse the output of a target web site for any sequence that matches the extraction pattern. Any output of the target web site that matches the extraction pattern can be returned as rendered code. If the extraction pattern is properly developed, the matching code will contain useful product information and not extraneous undesired portions of the target web site output. Dworkin and Gregory do not teach or suggest this aspect of the claimed invention.

Applicants' claims are patentable over any combination of Dworkin and Gregory

Dworkin teaches a service for obtaining product information. The user tells the system the general type of product or service desired. Typically, this is accomplished by menu selection in Dworkin. In response to the user's choice, the system displays a template that gives various technical criteria for the product or service. By filling in one or more spaces on this template, the user can tell the system the criteria of interest. The system then searches a database for products

that fulfill these criteria. Then, the system displays to the user the general information about the products or services rendered.

Even if Dworkin's menu selections were equated to the values that are used as **extraction parameters** in the claimed invention, Dworkin still does not teach or suggest the claimed invention because Dworkin uses such values to query a data source that, by its very nature, has a known output data format. Therefore, Dworkin does not have any need or motivation to develop Applicants' **extraction patterns**. Dworkin merely uses the menu selections and associated input to query a database for a result and then displays the result. In contrast, Applicants use the code generated by a plurality of web sites with initially unknown output data format in order to obtain data of interest.

Applicants method is advantageous over Dworkin because, unlike Dworkin, no database of information of interest has to be maintained in Applicants' method. However, unlike Dworkin, Applicants must overcome the problem of how to obtain useful information in a uniform, comprehensible manner from web sites having unknown output data formats. Applicants have addressed this problem using novel extraction patterns. Each extraction pattern is customized (developed) for each target web site so that exact, specific, information is retrieved from the respective web sites. In one example, there are two extraction patterns for each target web site. The first respective extraction pattern is customized (developed) so that it obtains the title of a book from a target web site (when a user passes a value for the book title) and the second respective extraction pattern is customized so that it obtains the price of the book from the target web site. Each of the plurality of web sites (*e.g.*, Barnes and Noble, Amazon, *etc.*) encodes such information in a different way. Thus, unique extraction patterns are developed for each of these target web sites.

Gregory does not remedy the failure of Dworkin to teach or suggest Applicants' claimed invention. Like the instant application and Dworkin, Gregory can provide product information from a plurality of sources. However, Gregory operates by dividing the task in a cooperative, highly structured manner between a commerce server and a plurality of merchant servers. The commerce server provides a limited amount of information about products that were supplied from partner merchants. When the user requests more detailed information about a particular product, process control is passed from the commerce server directly to a particular merchant

server. However, at no time does Gregory develop extraction patterns based on the output of target web sites. Rather, Gregory ultimately passes control to a single merchant server without ever scanning the output of the merchant server using an extraction pattern.

For the above reasons, the pending claims are patentable over any combination of Dworkin and Gregory. Therefore, Applicants respectfully request that the rejection to claims 1-21 be withdrawn.

THE CLAIMS AS AMENDED ARE PATENTABLE OVER HERZ

In the previous office action, dated January 16, 2003, the Examiner rejected claims 1-21 under 35 U.S.C. 102(b) as being anticipated by United States Patent 5,754,938 to Herz *et al.* (hereinafter "Herz"). Because of the amendments to the pending claims, Applicants would like to discuss why the pending claims, as amended, remain patentable over Herz. Herz does not disclose developing a description of data of interest for each web site in a plurality of web sites, based on the output generated by the plurality of web sites, each respective description of data of interest specifying an address for a corresponding web site in the plurality of web sites and each respective description of data of interest including an extraction pattern that is capable of extracting user specified information from a corresponding web site. Rather, Herz matches user profiles to the profile associated with objects. Further, Herz returns whole objects (news clippings, documents, *etc.*) whereas Applicants' extraction patterns ensure that only very specific portions of the output of the web site are captured and displayed.

THE CLAIMS AS AMENDED ARE PATENTABLE OVER KIM IN VIEW OF CHELLIAH

In the office action dated June 28, 2002, the Examiner rejected the claims under 35 U.S.C. 103 as being unpatentable over the teachings of Kim "WebData.com Debuts Online Shopping Price Comparisons" (hereinafter "Kim") in view of United States Patent No. 5,710,887 to Chelliah *et al.* (hereinafter "Chelliah"). In response to the June 28, 2002 office action,

Applicants amended the claims to recite the simultaneous provision of product information when the product information includes product information for at least two sources in a plurality of sources. In fact, Applicants invention is not limited to simultaneous provision of product information when such information is available from multiple sources. As indicated in new claim 25, the data of interest can be provided incrementally as the data of interest is extracted from the web sites. Thus, in the present amendment, this "simultaneous" limitation has been removed from the independent claims.

Applicants claims are fully patentable over Kim and Chelliah, either alone or in combination, without the "simultaneous" limitation, for at least the following reasons.

Chelliah discloses an electronic shopping mall with a plurality of storefronts. However, Chelliah does not develop extraction patterns because the product information is stored in a customized product database (e.g., database 116, Fig. 5). As such, the product information is outputted in a known format and there is no need for the development of Applicants' extraction patterns in Chelliah. Further, Chelliah does not query a plurality of web sites as indicated in Applicants' claims because the user enters a particular store in Chelliah by selecting a single Electronic Storefront 14 (See, for example, Chelliah at column 6, lines 37-43).

Kim discloses the database search engine "WebData.com", which was developed by ExperTelligence, Inc. Exhibit A is a September 30, 2000 ExpertTelligence, Inc. Annual Report to Stockholders. Page 5 of the report states that WebData.com consists of 5,000 high quality cataloged databases. Exhibit B is an ExperTelligence, Inc. Press Release that states that Webdata.com employs Expertelligence's Agent3W technology to collect, organize and build relationships between searchable databases. Exhibit C is a Lawyerware survey that states that Webdata collects databases from the Internet into one place using an Agent3W spider and HTML parsing technology. Exhibit D is a October/November 2001 Imprints Circular from the Central Connecticut Chapter of the Society for Technical Communication. Like Exhibit C, page 6 of Exhibit D indicates that Webdata.com uses an Agent3W spider and HTML parsing technology to search only on-line databases.

Exhibit E is a portion of the documentation for build 67 of WebBase. Page 5 of the September 30, 2000 ExpertTelligence, Inc. Annual Report to Stockholders (Exhibit A) states that WebBase is a software product for providing easy access to database information over the

Internet. From this, it is clear that the WebData.com portal described in Kim is driven using WebBase's Agent3W.

Exhibit E includes a table of contents with the entry "What is Agent3w" (Exhibit E, page 1). A detailed description of Agent3W from the documentation for build 67 of WebBase begins on page 4 of Exhibit E. The detailed description in Exhibit E shows that Agent3W obtains information from web sites using standard HTTP "get" and "post" commands. Thus, unlike Applicants' claimed invention, Agent3W does not parse the output of target web sites looking for matches to extraction patterns.

CONCLUSION

In light of the above amendments and remarks, the Applicants respectfully request that the Examiner reconsider this application with a view towards allowance. The Examiner is invited to call the undersigned attorney at (650) 493-4935, if a telephone call could help resolve any remaining items.

Respectfully submitted,

Date: September 3, 2003

Gary S. Williams

31,066

(Reg. No.)

PENNIE & EDMONDS LLP
3300 Hillview Avenue
Palo Alto, California 94304
(650) 493-4935